

**METHODS AND SYSTEMS FOR THE IDENTIFICATION OF RNA  
REGULATORY SEQUENCES AND COMPOUNDS THAT MODULATE THEIR  
FUNCTION**

**CROSS REFERENCE TO RELATED APPLICATION**

The present application claims priority to U.S. Provisional Application No.  
60/441,028, filed January 17, 2003.

5

**FIELD OF THE INVENTION**

The present invention relates to the identification of RNA regulatory sequences  
and compounds that modulate gene expression at the post-transcriptional level. More  
specifically, the invention relates to the screening for RNA sequences able to inhibit the  
translation of a reporter mRNA and for compounds able to reverse the inhibition of

10

translation.

**BACKGROUND OF THE INVENTION**

Gene expression is controlled at many different steps in the pathway from DNA to  
RNA to protein. Because aberrant gene expression can lead to a disease state, such as  
cancer, genes must be tightly regulated to ensure they are expressed at the correct time,  
place and level. While most efforts have been aimed at understanding transcriptional  
regulation of gene expression (*i.e.*, DNA to RNA) and its contribution to disease,  
regulation at other levels such as mRNA translation (*i.e.*, RNA to protein) or RNA  
stability remains less well understood. It is only recently that research into post-

15

transcriptional mechanisms of gene expression has uncovered that regulation of mRNA translation, or translational control, is a critical checkpoint in gene expression linked to a variety of disease processes (Cazzola and Skoda, Blood 95: 3280-3288, 2000).

Translational control occurs in virtually all cell types and species where it contributes to such diverse processes as cell-cycle control, learning and plasticity in neurons, and red blood cell differentiation, among many others. Because translational control enables a cell to increase the concentration of a protein very rapidly, this mechanism of control is especially suited for the regulation of genes that are involved in cell proliferation and damage control. Regulation of gene expression at the level of mRNA translation is also particularly important in cellular responses to development or environmental stimuli – such as nutrient levels, cytokines, hormones, and temperature shifts, as well as environmental stresses – such as hypoxia, hypocalcemia, viral infection, and tissue injury. Translational control can be either global, affecting all the mRNAs in a cell, or specific to a single or subset of mRNAs.

The typical mRNA contains a 5' cap, a 5' untranslated region upstream of a start codon (5' UTR), an open reading frame, also referred to as coding sequence, that encodes a functional protein, a 3' untranslated region (3' UTR) downstream of the termination codon, and a poly(A) tail. The key mediators of translational control are typically found in the 5' and 3' untranslated regions of mRNA transcripts, although the possibility of regulatory sequences mapping even to the coding sequence itself cannot be excluded. Much like the linear array of amino acids in proteins, these single-stranded regions of RNA can fold into complex three-dimensional structures consisting of local motifs such as hairpins, stem-loops, bulges, pseudoknots, guanosine quartets, and turns (for reviews

see Moore, *Ann. Rev. Biochem.* 68:287-300, 1999; Gallego and Varani, *Acc.Chem. Res.* 34:836-843, 2001). Through interactions with regulatory proteins, such structures can be critical to the activity of the nucleic acid and dramatically affect the regulation of mRNA translation.

5           Because the sequences of an mRNA often contain critical regulatory elements which influence translational efficiency, compounds that are able to modulate the effect of the regulatory RNA sequence would be highly useful in therapeutic applications that seek to up- or downregulate the expression of a gene. Current approaches for blocking the function of target nucleic acids include the use of duplex-forming antisense  
10   oligonucleotides (Bennett and Cowser, *Biochem. Biophys. Acta* 1489 (1): 19-30, 1999), peptide nucleic acids ("PNA"; Gambari, *Curr. Pharm. Des.* 7 (17): 1839-1862, 2001; Nielsen, *Curr. Opin. Struct. Biol.* 9 (3): 353-357, 1999; Nielsen, *Curr. Opin. Biotechnol.* 10 (1): 71-75, 1999) and locked nucleic acid ("LNA"; Braasch & Corey, *Chem. Biol.* 8 (1): 1-7, 2001; Arzumanov *et al.*, *Biochemistry* 40 (48): 14645-14654, 2001), which bind to  
15   nucleic acids *via* Watson-Crick base-pairing. However, the dependence on the native three-dimensional structural motifs of single-stranded stretches for regulatory functions can preclude the use of general, simple-to-use, sequence-specific recognition rules to design complementary agents that bind to these motifs.

          Previous efforts to identify compounds or agents that recognize regulatory RNA  
20   elements have primarily focused on characterizing regulatory proteins that bind to a particular regulatory mRNA sequence, and on elucidating molecular mechanisms by which the protein-mRNA complex exerts its effect on translational control before identifying potential modulators. A major disadvantage of such approaches is the lengthy

and laborious procedure required to isolate and identify proteins that bind to specific mRNA regulatory sequences. In addition to isolating the proteins that bind to regulatory mRNA sequences, these approaches have also either required the labeling of particular proteins or RNAs or depended on the linkage of the RNA regulatory sequence to a reporter, or a combination thereof. All these are time-consuming and laborious procedures that require a series of complex laboratory manipulations and often deliver false positive results. There is thus a need for a simplified method to identify modulators of translational control of gene expression that eliminates the requirement for a series of intermediate steps and yields a direct functional readout.

## SUMMARY OF THE INVENTION

The present invention provides a non-cell based method of screening for and/or identifying an RNA regulatory element. The method includes combining a translation extract; an RNA test sequence; and a reporter mRNA under suitable conditions for translation of the reporter mRNA. The method further includes measuring the effect of the RNA test sequence on the translation of the reporter mRNA, wherein a test sequence that modifies the translation of the reporter mRNA includes an RNA regulatory element.

In one embodiment, a test sequence which inhibits translation of the reporter mRNA, as compared to in the absence of the test sequence, includes an RNA regulatory element. In another embodiment, a test sequence which increases the translation of the reporter mRNA, as compared to in the absence of the test sequence, includes an RNA regulatory element.

The present invention further provides a non-cell based method of screening for and/or identifying at least one test compound, which modulates the ability of an RNA sequence to regulate translation of a reporter mRNA. The method includes: providing an RNA sequence, which regulates translation of a reporter mRNA; and combining the RNA  
5 sequence with a translation extract; the reporter mRNA; and at least one test compound under conditions suitable for translation of the reporter mRNA. The method further includes measuring the effect of the at least one test compound on the ability of the RNA sequence to regulate the translation of the reporter mRNA. For example, the RNA regulatory sequence can inhibit the translation of the reporter mRNA. In this instance, the  
10 method can be used to assess whether a particular test compound(s) reverses the inhibition, as measured by an increase in translation of the reporter mRNA.

Another aspect of the invention relates to an *in vitro* translation system. The system includes a cytoplasmic translation extract; an RNA regulatory sequence; and a reporter mRNA; wherein the RNA regulatory sequence modifies the translation of the  
15 reporter mRNA. The system can be used in a screening method according to the present invention for identifying a test compound that modulates the ability of the RNA regulatory sequence to regulate translation of the reporter mRNA. In particular, in this method, a test compound is introduced into the system and the extent of modulation of translation of the reporter mRNA is determined.

20 A further aspect of the present invention relates to an *in vitro* translation system that includes a cytoplasmic translation extract; an RNA regulatory sequence; and a reporter mRNA; wherein the RNA regulatory sequence inhibits translation of the reporter mRNA. This system can be used in a screening method provided by the present invention

for identifying a test compound which is capable of reversing the inhibition of translation. In particular, in this method, a test compound is introduced into the system; and the extent of reverse inhibition of translation of the reporter mRNA is assessed.

Also provided by the invention is a test compound identified by a screening  
5 method of the present invention and a use therefore. For example, a test compound identified in a screening method of the present invention can be used in the manufacture of a medicine for modulating the expression of a gene including the RNA sequence. For example, the RNA sequence can be harbored within a gene involved in pathogenesis and/or pathophysiology. The expression of the gene may be aberrant in a disease state or  
10 may cause the survival and/or progression of a pathogenic organism.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

**Figure 1. (A)** This figure shows the proposed mechanism of reverse inhibition of translation. The translation initiation process is a key step in the regulation of gene expression. RNA elements that mediate this regulation (regulatory RNA), when added  
15 exogenously to an *in vitro* translation system, can interact with any number of a large array of general translation components, such as the 40S ribosomal subunit (40S sphere), and inhibit gene expression. Test compounds (cylinder) that interact with the RNA regulatory element can reverse the inhibition of gene expression and modulate the expression of genes that harbor the regulatory RNA element. **(B)** This figure represents  
20 the proposed mechanism of reverse inhibition as an enzymatic reaction whose components are in thermodynamic equilibrium. In the absence of inhibitory RNA (I, inhibitor), the ribosome (E, enzyme) translates the mRNA (S, substrate) to generate the

reporter enzyme (P, product). In the presence of I, E is sequestered and P formation is inhibited. However, the addition of a test compound (C, compound) that prevents the interaction between E and I allows free E to translate S for the production of P.

Translation of S by E in the presence of C is herein referred to as Reverse Inhibition.

5       **Figure 2. (A)** This figure contains sequence coding for nucleotides 18 to 356 of the internal ribosome entry site (IRES) RNA of Hepatitis C Virus (HCV) and corresponds to SEQ ID NO. 1. HCV-IRES RNA (genotype 2b) is inserted at the BamHI restriction endonuclease site (underlining) of the cloning vector pGEM-3 (Promega Corp., Madison, WI.). **(B)** This figure shows the secondary structure of wild-type HCV 5' UTR RNA as  
10       predicted by *in silico* algorithms and verified by experimental techniques (Honda *et al.*, J. Virol. 73 (2): 1165-1174, 1999). The information from such a representation can be used to divide the sequence into fragments for the determination of a minimal sequence that is able to mediate inhibition of reporter translation. Non-structured, inter-domain portions of the sequence are used as primer binding sites for fragment generation by PCR. The  
15       IRES contains domains II, III and IV (nt. 44 –354). The 40s ribosome subunit is proposed to interact with subdomain III<sub>d</sub>, while eIF3, part of the translation initiation complex, is proposed to interact with subdomains III<sub>b</sub> and III<sub>c</sub>.

**Figure 3.** This figure shows the ability of the different fragments of the HCV-IRES UTR to inhibit translation of the reporter gene. An *in vitro* translation system was  
20       preincubated with the RNA fragments before addition of the reporter mRNA construct encoding firefly luciferase (Fy Luc). The inhibitory effects of each of the fragments on general translation were determined by monitoring luciferase activity upon addition of substrate. Note that the majority of inhibitory activity resides in domain III (solid bar);

accordingly, domain III was identified as the minimal RNA element required for efficient translation inhibition and applied to the reverse-inhibition assay. Domain I not analyzed due to lack of any predicted secondary structure.

**Figure 4.** This figure shows the components of an exemplary reporter DNA construct used to produce reporter mRNA for use in the identification of compounds with the ability to reverse inhibition of translation. This reporter uses the open reading frame (ORF) of firefly luciferase fused to a consensus T7 promoter sequence at the NcoI and BglII restriction endonuclease cleavage sites and is devoid of any predicted RNA regulatory elements. The linearized template used for run-off transcription of mRNA is generated by restriction endonuclease digestion at the BamHI site. The resulting transcription product is purified by precipitation with LiCl and is used without further modification.

**Figure 5.** This figure shows the ability of test compound PTC-0099870 to reverse the inhibition activity of domain III of the HCV-IRES UTR. A cell-free translation extract (RRL) was preincubated with the RNA fragments and test compound before addition of the reporter mRNA construct (firefly luciferase). The data shown represents the reverse inhibitory effects of the test compound at increasing concentrations (0-28  $\mu$ M) on general translation as determined by monitoring luciferase activity upon addition of substrate. Nearly complete reverse-inhibition, whereby expression levels are equivalent to the translation reaction performed in the absence of IRES III RNA fragment, is achieved at a compound concentration of 28  $\mu$ M.

#### **DETAILED DESCRIPTION OF THE INVENTION**



Various publications or patents are referred to in parentheses throughout this application to describe the state of the art to which the invention pertains. Each of these publications or patents is incorporated by reference herein.

The present invention relates to methods for screening and identifying RNA regulatory sequences and test compounds that modulate the expression of genes at post-transcriptional events. The terms "RNA regulatory element", "RNA regulatory sequence" or "RNA element" are used herein along with "UTR" and "UTR regulatory element," to denote those RNA sequences – both RNA only and/or protein-RNA complexes – that influence or regulate the translation machinery, be it positively by upregulating translation efficiency, or negatively by downregulating or inhibiting translation, regardless of where in the transcript they are located, *i.e.* in untranslated regions or in the coding sequence. These RNA regulatory elements, when introduced exogenously to a cell-free *in vitro* translation system containing reporter mRNA and cytoplasmic extract which may additionally be supplemented with amino acids, tRNA, hemin, creatine kinase, KOAc, Mg(OAc)<sub>2</sub>, and creatine phosphate, can act as antagonists of gene expression through direct competition for essential components of the general translational machinery, or through recruitment of regulatory proteins that interact with essential components of the general translation machinery. Significantly, the ability of the RNA regulatory elements to inhibit the translation of the reporter mRNA when introduced exogenously to a cell-free *in vitro* translation system does not depend on whether the endogenous RNA regulatory element functions to decrease or increase translation of a corresponding coding sequence. Thus, the present invention allows the identification of both positive and

negative RNA regulatory elements, as well as compounds that modulate the effects of both positive and negative RNA regulatory elements on the translational machinery.

Thus in a first aspect, the present invention allows for the speedy identification of novel RNA regulatory elements involved in translational control. By the systems and  
5 methods of the invention, any RNA test sequence of interest, or a fragment thereof, can be quickly and conveniently assayed for its ability to inhibit translation of a reporter mRNA. For example, in one embodiment, a suitable test sequence corresponds to a sequence from the 5' UTR or 3' UTR of an mRNA of a target gene of interest. In another embodiment, a suitable test sequence corresponds to a sequence from the coding region of an mRNA of a  
10 target gene of interest. In a further embodiments, the test sequence is from an mRNA of a gene involved in pathogenesis and/or pathophysiology, as will be described in further detail below.

The method according to the present invention for screening for and/or identifying an RNA regulatory element includes combining (i) a translation extract; (ii) an RNA test  
15 sequence and (iii) a reporter mRNA under suitable conditions for translation of the reporter mRNA. Once combined, the effect of the test sequence on the translation of the reporter mRNA is measured, wherein a test sequence that modifies the translation of the reporter mRNA includes an RNA regulatory element.

In a preferred embodiment, the combining step includes preincubating the  
20 translation extract with the test sequence; and then combining the preincubated extract with the reporter RNA. In another embodiment, the combining step includes preincubating the translation extract with the reporter mRNA; and subsequently

combining the preincubated extract with the test sequence. The measuring step can include detecting a signal resulting from translation of the reporter mRNA in the presence of the test sequence and comparing it to the signal resulting from translation of the reporter mRNA in the absence of the test sequence.

5        In one embodiment, a test sequence which inhibits the translation of the reporter mRNA, as compared to in the absence of the test sequence, includes an RNA regulatory element. In another embodiment, a test sequence which increases the translation of the reporter mRNA, as compared to in the absence of the test sequence, includes an RNA regulatory element.

10        In a further aspect, the instant invention allows for the identification of compounds or agents that modulate the regulatory activity of an RNA regulatory sequence of interest, as measured by a difference, such as an increase, in translation of the reporter mRNA, as compared to in the absence of the test compound. In particular, the invention provides a method of screening for and/or identifying at least one test compound which  
15        modulates the ability of the RNA regulatory sequence to regulate translation of a reporter RNA. The method includes providing an RNA sequence which regulates translation of a reporter mRNA; and combining the RNA sequence with: (i) a translation extract; (ii) the reporter mRNA; and (iii) the at least one test compound under suitable conditions for translation of the reporter mRNA. The method also includes the step of measuring the  
20        effect of the at least one test compound on the ability of the RNA sequence to regulate the translation of the reporter mRNA. The measuring step can include detecting a signal resulting from translation of the reporter mRNA in the presence of the test compound, and

comparing it to the signal resulting from translation of the reporter mRNA in the absence of the test compound.

In one embodiment of the methods to screen test compounds, the step of providing the RNA sequence includes contacting the RNA sequence with an *in vitro* translation  
5 system and assessing translational modification of the reporter mRNA when the reporter mRNA is introduced into the contacted system. Thus, an RNA test sequence which includes an RNA regulatory element and influences and/or regulates translation of the reporter mRNA, as compared to in the absence of the RNA test sequence, can be employed to screen for test compounds which can modulate this regulatory activity. The  
10 RNA sequence employed can include an RNA regulatory element identified by a method of the present invention, or can include an already known RNA regulatory element

In any event, a test compound(s) can be screened by combining the RNA sequence with a translation extract; the reporter mRNA; and the at least one test compound, wherein the RNA sequence modifies translation of the reporter mRNA. In a desired  
15 embodiment, the combining step includes preincubating the translation extract with the RNA sequence and the test compound; and combining the preincubated extract with the reporter mRNA.

In another embodiment, the method assesses whether a test compound(s) inhibits the interaction between the RNA sequence and one or more components in the translation  
20 extract. For example, the RNA sequence can increase or, alternatively, decrease the translation of the reporter mRNA by interacting with one or more components of the translation machinery, and a test compound may influence and/or modify this interaction.

In one embodiment, the RNA sequence employed in the screen for target compounds inhibits the translation of the reporter mRNA. In this instance, the method can be used to assess whether or not a test compound reverses the inhibition, as measured by an increase in translation of the reporter mRNA. This is illustrated in Figure 1A, a description for which is in the Brief Description of the Drawings.

Test compounds that inhibit the interaction between the exogenously added RNA regulatory elements and one or more components of the general translational machinery relieve the antagonistic effect of the RNA regulatory elements on the translation of the reporter mRNA and cause an increase in gene expression, *i.e.*, translation of the reporter mRNA. The increase in the level of gene expression upon addition of test compounds is referred to herein as “reverse inhibition”, and forms a basis for the identification of molecules that modulate gene expression through direct interactions with the exogenously added RNA regulatory element. The proposed mechanism of reverse inhibition is shown in Figures 1A and 1B, descriptions for which are in the Brief Description of the Drawings.

Test compounds identified using the methods and systems of the present invention are useful as therapeutics for modulating the expression of genes that harbor the RNA elements(s) and whose expression is aberrant in the disease state (Mendell and Dietz, Cell 107: 411-414, 2001; Keene and Tenenbaum, Mol. Cell 9:1161-1167, 2002), or whose expression causes the survival and/or progression of a pathogenic organism. Furthermore, said compounds are useful as molecular tools for regulating the expression of recombinant proteins expressed from constructs engineered to contain the RNA elements(s).

### Description of Advantages to the Experimental Setup

In preferred embodiments of the invention, the RNA test sequence of interest, or a regulatory fragment thereof, is added to a translation extract and the combination of the translation extract and RNA fragment are contacted with the reporter mRNA to assess inhibition of translation of the reporter. Once the RNA test sequence or fragment thereof is found to inhibit the translation of the reporter mRNA, the system including the inhibitory RNA test sequence, reporter mRNA and translation extract, is contacted with a library of test compounds to assay for the reverse-inhibition of translation of the reporter mRNA. The invention provides several significant advantages over previous approaches to identifying modulators of gene expression that target post-transcriptional regulation. One of the advantages over previous approaches is the present invention's ability to target both known and unknown RNA regulatory elements. Typical drug discovery systems (*i.e.*, screens) for identifying modulators of gene expression that target post-transcriptional regulation require the identification and characterization of the RNA regulatory element and its interacting molecule(s). In all cases, the exact nature of the RNA regulatory element must first be known or determined before proceeding with the establishment of a screen to identify modulators. The screening system for test compounds set forth herein requires only ability of the RNA test sequence or fragment thereof to inhibit translation when added exogenously to an *in vitro* translation system. Thus, the present invention is capable of combining a rapid screening system for identifying compounds that interact with RNA regulatory elements and modulate gene expression, with a system of determining whether an RNA test sequence is involved in translational control. Because the two screening functions can be performed simultaneously with the systems and

methods of the present invention, the speed and efficiency of therapeutic drug discovery are further increased.

Another advantage of the approach of the present invention is the specificity of its signal output. The present approach detects specific interactions between the test  
5 compound and target RNA through an increase in signal from the reporter gene expression. Expression of the reporter mRNA increases when a compound interacts with the RNA test sequence so as to re-activate the components of the translational machinery and to allow translation of the reporter. In contrast to many other approaches, the present invention detects neither reporter gene antagonists, nor reporter enzyme antagonists, both  
10 of which yield the typical false positive results in other assays that monitor a decrease in signal from the reporter gene expression (*i.e.*, inhibition). For example, most assays are cell-based and feature a reporter enzyme coding sequence attached to a predetermined regulatory sequence of interest. The readout of such assays consists of a change in the expression of the reporter sequence upon addition of a test compound that interacts with  
15 the regulatory sequence of interest. However, a test compound that affects the enzymatic activity of the reporter instead of modulating the activity of the regulatory sequence generates a false positive or false negative signal in such assays. By contrast, in the systems and methods of the present invention, a test compound that inhibits the enzymatic activity of the reporter would correctly generate a negative result (decrease in signal or  
20 absence thereof) and be excluded from the target group of compounds that specifically interact with the regulatory RNA fragment.

In addition, most known assays detect cytotoxic agents and general inhibitors of translation as false positives in inhibition assays (including those not interacting with

RNA), whereas in the present invention, such general inhibitors of translation are precluded from generating positive signals because they would also inhibit the translation of the reporter mRNA. Furthermore, interactions between the test compound and factors of the general translation machinery involved in inhibition would generate a negative  
5 result if they in turn have inhibitory effects on translation, leading to the appropriate exclusion of such nonspecific (and thereby toxic) agents from the target group of compounds. Test compounds generate positive results in the systems and methods of the present invention if they interfere with the interaction between the RNA test sequence of interest and a component of the general translation machinery. Thus, test compounds  
10 identified by the present invention target specifically the RNA test sequence of interest, leading to a dramatic reduction in the number of artificial results obtained in prior approaches. This advantageous feature of the present invention makes it particularly well suited to high-throughput screening for RNA-interacting molecules that modulate gene expression.

15 Finally, the systems and methods of the present invention can be performed under controlled and tunable conditions. Specifically, the reporter gene expression has a fixed window for reverse-inhibition, or reference range, that is determined through translation of reporter mRNA without the RNA regulatory element present. Having a fixed window for activation (*i.e.*, 100% reverse-inhibition) allows one to rank-order active compounds  
20 based on their levels of reverse inhibition. Furthermore, by simply increasing or decreasing the amount of inhibitory RNA, one is able to adjust the window of reverse-inhibition and, thereby, "tune" the stringency of the screen (*i.e.*, the ability of the screen to observe positive signals). Thus, the specificity of the invention allows for the rapid



generation of highly specific positive signals with rank-ordering ability, which is also highly desirable in high-throughput screening for RNA-interacting molecules that modulate gene expression.

### **Description of RNA Test Sequences and/or Candidate RNA Regulatory Elements**

5        Examples of RNA regulatory elements from 5' UTRs, which are well known in the art include Iron response element (IRE), Internal ribosome entry site (IRES), upstream open reading frame (uORF), Male specific lethal element (MSL-2), G-quartet element, and 5'-terminal oligopyrimidine tract (TOP). See, for example, *Translational control of gene expression*, Sonenberg, Hershey, and Mathews, eds., CSHL Press, 2000. Examples  
10    of known 3' UTR regulatory elements include AU-rich elements (AREs), ARE enhancers, Selenocysteine insertion sequence (SECIS), Histone stem loop, Cytoplasmic polyadenylation elements (CPEs), Nanos translational control element, Amyloid precursor protein element (APP), Translational regulation element (TGE)/direct repeat element (DRE), Bruno element (BRE), 15-lipoxygenase differentiation control element (15-LOX-DICE), and G-quartet element (Keene and Tenenbaum, Mol Cell 9:1161-1167, 2002).  
15

      In a preferred embodiment, known regulatory RNA sequences for use in the systems of the present invention include the internal ribosome entry sites (IRES), which are among the best characterized 5' UTR-based *cis*-elements of post-transcriptional gene expression control. IRES elements facilitate cap-independent translation initiation by  
20    recruiting ribosomes directly to the mRNA start codon, are commonly located in the 3' region of 5' UTR, and are frequently composed of several discrete sequences. IRESes do not share significant sequence homology, but do form distinct RNA tertiary structures.

Some IRESes contain sequences complementary to 18S RNA and form stable complexes with the 40S ribosomal subunit and initiate assembly of a translationally competent complex. A classic example of an IRES is the internal ribosome entry site from Hepatitis C virus. Most known IRESes require protein co-factors for activity. More than 10 IRES  
5 *trans*-acting factors (ITAFs) have been identified so far. In addition, all canonical translation initiation factors, with the sole exception of 5' end cap-binding eIF4E, have been shown to participate in IRES-mediated translation initiation (reviewed in Vagner *et al.*, EMBO reports 2:893-898, 2001; *Translational control of gene expression*, Sonenberg, Hershey, and Mathews, eds., CSHL Press, 2000).

10 In another preferred embodiment, known regulatory RNA sequences for use in the systems of the present invention are AU-rich elements (AREs). AU-rich elements are the most extensively studied 3' UTR-based regulatory signals. AREs are the primary determinant of mRNA stability and one of the key determinants of mRNA translation initiation efficiency. A typical ARE is 50 to 150 nucleotides long and contains 3 to 6  
15 copies of an AU<sub>n</sub>A sequence (where n = 3, 4, or 5) embedded in a generally A/U-enriched RNA region. The AU<sub>n</sub>A sequence be scattered within the region or can stagger or even overlap (Chen *et al.*, TIBS 20:465-470, 1995; Wiklund *et al.*, JBC 277:40462-40471, 2002; Tholanikunnel and Malborn, JBC 272:11471-11478, 1997; Worthington *et al.*, JBC Sep 24, 2002). The activity of certain AU-rich elements in promoting mRNA degradation  
20 is enhanced in the presence of distal uridine-rich sequences. These U-rich elements do not affect mRNA stability when present alone and thus that have been termed "ARE enhancers" (Chen *et al.*, Mol. Cell. Biol. 14:416-426, 1994).

Most AREs function in mRNA decay regulation and/or translation initiation regulation by interacting with specific ARE-binding proteins (AUBPs). AUBP functional properties determine ARE involvement in one or both pathways. For example, ELAV/HuR binding to *c-fos* ARE inhibits *c-fos* mRNA decay (Brennan and Steitz, Cell Mol Life Sci. 58:266-277, 2001), association of tristetraprolin with TNF $\alpha$  ARE dramatically enhances TNF $\alpha$  mRNA hydrolysis (Carballo *et al.*, Science 281:1001-1005, 1998), whereas interaction of TIA-1 with the TNF $\alpha$  ARE does not alter the TNF $\alpha$  mRNA stability but inhibits TNF $\alpha$  translation (Piecyk *et al.*, EMBO J. 19:4154-4163, 2000). The competition of multiple AUBPs for the limited set of AUBP-binding sites in an ARE and the resulting "ARE proteome" determines the ARE regulatory output (Chen *et al.*, Cell 107:451-464, 2001; Mukherjee *et al.*, EMBO J. 21:165-174, 2002). Furthermore, the effects of AREs depends on ongoing translation (Curatola *et al.*, Mol. Cell. Biol. 15:6331-6340, 1995; Chen *et al.*, Mol. Cell. Biol. 15:5777-5788, 1995; Koeller *et al.*, PNAS 88:7778-7782, 1991; Savant-Bhonsale *et al.*, Genes Dev. 6:1927-1939, 1992; Aharon and Schneider, Mol. Cell. Biol. 13:1971-1980, 1993). It is not clear how a 3' UTR-localized element can affect translation initiation -- a process that takes place in the 5' UTR. One plausible explanation comes from recent work showing that most or all cytoplasmic mRNAs are circularized *via* eIF4F - poly(A)-binding protein (PABP) interaction; this interaction connects the two UTRs and can bring AREs in the 3' UTR into close proximity to the translation initiation site (Wells *et al.*, Mol. Cell. 2:135-140, 1998). Thus, the translation machinery, in addition to its role in translating mRNA, can also serve as a destabilizing/ ribonuclease-recruiting or a stabilizing/AUBPs-removing entity.

The methods and systems of the present invention can be applied to any target gene of interest. Specifically, the invention contemplates the identification of regulatory RNA sequences and the modulation by compounds identified by the instant methods of genes harboring these sequences that are involved in pathogenesis and pathophysiology.

5 Thus, for example, genes involved in carcinogenesis are suitable candidates for the methods and systems of the present invention. Such genes comprise oncogenes, *i.e.*, genes associated with the stimulation of cell division, such as, but not limited to, genes coding for growth factors or receptors for growth factors, *e.g.*, PDGF (brain and breast cancer), erb-B receptor for epidermal growth factor (brain and breast cancer), erb-B2  
10 receptor for growth factor (breast, salivary, and ovarian cancers), RET growth factor receptor (thyroid cancer), Ki-ras activated by active growth factor receptor proteins (lung, ovarian, colon and pancreatic cancer), N-ras activated by active growth factor receptor proteins (leukemia's), c-src, a protein kinase that becomes overactive in phosphorylation of target proteins, transcription factors that activate growth promoting genes, such as c-  
15 myc which activates transcription of growth stimulation genes (leukemia, breast, stomach, and lung cancer), N-myc (nerve and brain cancer), L-myc (lung cancer), c-jun and c-fos, Bcl-2 which blocks cell suicide (lymphoma), Bcl-1 which codes for cyclin D1, a stimulatory protein of the cell cycle (breast, neck, head cancers), MDM2 which codes for antagonist of p53 (sarcomas).

20 Additionally, tumor suppressor genes, such as APC (colon and stomach cancers), DPC4 which is involved in cell division inhibitory pathway (pancreatic cancer), NF-1 which inhibits a stimulatory (Ras) protein (brain, nerve, and leukemia), NF-2 (brain and nerve cancers), MTS1 which codes for p16 which inhibits cyclin D-dependent kinase

activity (many cancers), RB, a master brake on cell cycle (retinoblastoma, bone, bladder, lung, and breast cancer), p53 which halts cell cycle in G1 and induces cell suicide (many cancers), WT1 (Wilms tumor of the kidney), BRCA1 and 2 which function in repair of damage to DNA (breast and ovarian cancers), VHL (kidney cancer), telomerase which is involved in tumor cell immortality, thymidylate synthase and many more known to those of skill in the art. In addition, genes coding for cytokines or virokinases are also suitable targets for the methods and systems of the present invention. Both cytokines and virokinases are well characterized in the art and available, for instance, from the Cytokines Online Pathfinder Encyclopaedia (<http://www.copewithcytokines.de/cope.cgi>).

Genes involved in other pathophysiological processes, including viral genes (*i.e.* HIV, HCV and others) are either published in the literature and thus well known in the art and/or are easily identified by a person of skill. For example, if a gene involved in a particular disease process has not already been widely published, the National Cancer Institute's Cancer Genome Anatomy Project offers the Gene Ontology browser which classifies genes by molecular function, biological process, and cellular component (<http://cgap.nci.nih.gov/Genes/GeneInfo?ORG=Hs&CID=193852>), while the Human Gene Mutation database database can be searched either by disease, gene name or gene symbol (<http://archive.uwcm.ac.uk/uwcm/mg/hgmd/search.html>). Due to the wealth of information about genes and their involvement in physiological processes as well as their dysregulation in disease that is available to a person skilled of art, it is evident that the systems and methods of the present invention can be practiced on any gene and its corresponding RNA sequence.

From the above, it is readily apparent that the instant invention is not limited to the use of UTRs, *i.e.* the untranslated sequences from the 5' cap to the start codon in case of a 5' UTR, or from the stop codon to the polyA tail in the case of a 3' UTR, but encompasses all mRNA fragments, even those potentially located within the coding sequence, that are  
5 capable of inhibiting translation when added exogenously. Thus, the invention allows not only the identification of novel RNA fragments able to inhibit the translation of the reporter mRNA, but also facilitates the characterization of the minimal regulatory elements contained within RNA sequences shown to be involved in translational control.

Accordingly, in another preferred embodiment of the present invention, synthetic  
10 RNA sequences are screened for the presence of regulatory elements. Chemical synthesis of oligonucleotides is a process well known in the art. The chemical synthesis can be that of DNA sequences which are subsequently transcribed into RNA by run-off transcription described *infra*, or the synthesis can be of RNA directly. Synthetic RNA (or DNA) sequences can be produced by the random incorporation of all four natural nucleotides, as  
15 well as non-natural nucleotides known to those of skill in the art, at each coupling step of solid phase synthesis. In addition to the component bases, a number of reagents are used to assist in the formation of internucleotide bonds (*e.g.*, oxidation, capping, detritylation, and deprotection). Automated synthesis is performed on a solid support matrix that serves as a scaffold for the sequential chemical reactions. (See also, *Oligonucleotide*  
20 *synthesis: A practical approach*, Atkinson T. and Smith M., IRL Press, Oxford, United Kingdom, 1984). The regulatory RNA sequences identified by using the methods and system of the present invention to screen random synthetic RNA sequences can be incorporated into expression vectors and used to increase the expression of recombinant

proteins in both cell-based and *in vitro* translation systems. Thus, such RNA regulatory elements, although not necessarily occurring in nature, can be used to enhance the expression of recombinant protein products of interest in a variety of biotechnology applications.

## 5                                      **Description of Isolation of Regulatory RNA Sequences**

Identification and isolation of mRNA is well known to those of skill in the art. Thus, for example, the cDNAs obtained by reverse transcription of the mRNA obtained from any source, including viruses, pathogenic organisms, individual cells or whole tissue, can be generated by the use of primers that hybridize specifically to sequences in  
10 the polyA tail, thus resulting in a cDNA library. cDNA libraries of many sources are also commercially available. The region of interest may then be amplified by PCR for use in the systems and methods of the present invention to obtain a DNA copy of the mRNA.

As discussed *supra*, many RNA regulatory elements are present in the untranslated regions of the mRNA. Thus, in a preferred embodiment of the present invention, a  
15 therapeutic gene of interest (or genome in the case of many viruses) is identified and its UTR sequences are located. Identification of known UTRs can be conveniently performed by the use of bioinformatics, such as database mining from GENBANK, where sequences are annotated to delineate the coding portion from the non-coding portion of a gene. Alternatively, a known mRNA regulatory element may, for example, be selected  
20 from those made available by the European Bioinformatics Institute (<http://srs.ebi.ac.uk/srs6bin/cgi-bin/wgetz?page+Libinfo+id+513761K9vhs+lib+UTR>), the French Institute of Health and Medical Research (<http://www.rangueil.inserm.fr/>

5 IRESdatabase), the UTR home page, a specialized sequence collection, deprived from redundancy, of 5' and 3' UTR sequences from eukaryotic mRNAs (<http://bighost.area.ba.cnr.it/BIG/UTRHome>), a database that searches for similarity between a query sequence and 5' or 3' UTR sequences in UTRdb collections from Nucleic Acids Research available at <http://www3.oup.co.uk/nar/database/c/>, UTRSite (collection of functional sequence patterns located in 5' or 3' UTR sequences) available at [http://bighost.area.ba.cnr.it/srs6bin/wgetz?-e+\[UTRSITE-ID:\\*\]](http://bighost.area.ba.cnr.it/srs6bin/wgetz?-e+[UTRSITE-ID:*]), UTRScan (looks for UTR functional elements by searching through user submitted query sequences for the patterns defined in the UTRsite collection) available at <http://bighost.area.ba.cnr.it/BIG/UTRScan/>, as well as UTRBlast at <http://bighost.area.ba.cnr.it/BIG/Blast/BlastUTR.html>, which searches for similarity between a query sequence and 5' or 3' UTR sequences in UTRdb collections, and other similar public databases and publications.

Alternatively, if the UTR sequences of the gene of interest are unknown, they can be identified experimentally by methods well known to those of skill in the art. For instance, the gene of interest can be cloned from a cDNA library and the ends of the cDNA can be amplified by RACE (rapid amplification of cDNA ends) and sequenced. Rapid Amplification of 5' cDNA Ends (5'-RACE) is used to extend partial cDNA clones by amplifying the 5' sequences of the corresponding mRNAs. The technique requires knowledge of a small region of sequence within the partial cDNA clone. During PCR, the thermostable DNA polymerase is directed to the appropriate target RNA by a single primer derived from the region of known sequence; the second primer required for PCR is complementary to a general feature of the target -- in the case of 5'-RACE, to a homopolymeric tail added (*via* terminal transferase) to the 3' termini of cDNAs



transcribed from a preparation of mRNA. This synthetic tail provides a primer-binding site upstream of the unknown 5' sequence of the target mRNA. Rapid Amplification of 3' cDNA Ends (3'-RACE) reactions are used to isolate unknown 3' sequences or to map the 3' termini of mRNAs onto a gene sequence. A population of mRNAs is transcribed  
5 into cDNA with an adaptor-primer consisting at its 3' end of a poly(T) tract and at its 5' end of an arbitrary sequence of 30-40 nucleotides. Reverse transcription is usually followed by two successive PCRs. The first PCR reaction is primed by a gene-specific sense oligonucleotide and an antisense primer complementary to the arbitrary sequence in the (dT) adaptor-primer. If necessary, the products of the first PCR can be used as  
10 templates for a second "nested" PCR, which is primed by a gene-specific sense oligonucleotide internal to the first, and a second antisense oligonucleotide complementary to the central region of the (dT) adaptor-primer. The products of the amplification reaction are cloned into a plasmid vector for sequencing and subsequent manipulation.

15 Furthermore, the present invention allows the identification of RNA regulatory elements present within the coding, or translated, sequence of mRNAs of genes of interest. Because the coding sequence of a gene of interest is easily obtained, as it is typically the primary subject of publication in journals and public databases, the foregoing description of the isolation of an RNA sequence or region of interest applies, in simplified  
20 form, to RNA coding sequences as well.

Once the RNA sequence or region of interest has been isolated, it can then be analyzed for the presence of known regulatory sequences that can be synthesized for use in the systems and methods of the present invention. Search algorithms such as BLAST

allow one skilled in the art to identify sequences with homology to the regulatory sequences of interest. If no known regulatory sequences are present in the RNA regions, the entire RNA sequence, UTR element and/or fragments thereof can be synthesized for use in the present invention. The secondary structure of single-stranded RNA can be

5 analyzed *in silico* to identify regions of the RNA that are likely to fold into higher order structures and, thereby, perform a regulatory function using algorithms provided by, for example, M-FOLD, RNA structure (Zuker algorithm), Vienna RNA Package, RNA Secondary Structure Prediction (Belozersky Institute, Moscow, Russia) and ESSA, among many known algorithms. These higher order structures include harpins, stem-loops,

10 bulges, pseudoknots, guanosine quartets, and turns (for reviews see Moore, Ann. Rev. Biochem. 68: 287-300, 1999; Gallego and Varani, Acc. Chem. Res. 34, 836-843, 2001). An analysis of the higher order structures revealed by *in silico* predictions allows a person skilled in the art to design primers that are complementary to inter-domain regions of the sequence for use in PCR amplification and subsequent cloning for fragment generation.

15       Synthesis of RNA fragments can be performed by a variety of techniques known to those of skill in the art. Run-off transcription from a cloned DNA template can be performed by the use of *in vitro* transcription methods known to those of skill in the art (Srivastava *et al.*, Methods Mol Biol. 86:201-207, 1998); commercially available *in vitro* transcription kits, such as Megascript™ from Ambion (Austin, TX) which uses T7 RNA  
20 polymerase to transcribe RNA from DNA templates harboring a T7 promoter in high yields can also be used. To accomplish this, one skilled in the art can readily design primers that hybridize to the 5' and 3' end portions of the specific region(s) of interest from a full-length cDNA clone. Use of these primers for PCR amplification, and

subsequent cloning into a suitable vector downstream of a T7 promoter sequence, is readily performed by a person skilled in the art. Alternatively, solid-phase oligonucleotide synthesis can be performed using phosphoramidite chemistry, and custom synthesis of RNA oligos is commercially available (Dharmacon Research, Inc., Lafayette,  
5 CO).

In another preferred embodiment, the invention provides an RNA regulatory sequence in combination with its specific regulatory protein co-factor(s) for use in the screen. If the regulatory co-factor is known, those of skill in the art, using recombinant techniques, can readily prepare it. Alternatively, the RNA regulatory sequence of interest  
10 may be isolated as described *supra* and immobilized to a solid-phase support by methods known to those of skill in the art (*e.g.*, affinity-tag). Upon addition of cellular extract, the RNA regulatory sequence will retain the factor(s) specifically able to bind the immobilized RNA sequence, whether known or unknown, and all other factors will be removed. The combination of RNA regulatory sequence with its specific regulatory co-  
15 factor can then be eluted from the solid-phase support for use in the system of the present invention. Use of the regulatory RNA sequence in conjunction with its regulatory co-factor will allow for the identification of test compounds that interfere with the interaction between the RNA regulatory element (in combination with its specific regulatory co-factor) and components of the general translation machinery.

20

### **Description of Reporter mRNAs**

In another preferred embodiment of the present invention, a reporter mRNA construct is provided. Synthesis of reporter constructs is well known in the art and can be

performed as described above for the synthesis of the regulatory RNA sequences, *i.e.* T7 promoter-driven run-off transcription from linear DNA templates using T7 RNA polymerase. In a preferred embodiment of the present invention, the linear DNA templates contain a promoter sequence and a protein coding sequence to express the reporter protein. In one preferred embodiment, the reporter is constructed to contain no UTR regulatory element and, thereby, monitors general translation efficiency. The reporter mRNA can be selected from known reporters in the art, including but not limited to, firefly luciferase, renilla luciferase, click beetle luciferase, green fluorescent protein, yellow fluorescent protein, red fluorescent protein, cyan fluorescent protein, blue fluorescent protein, *beta*-galactosidase, *beta*-glucoronidase, *beta*-lactamase, chloramphenicol acetyltransferase, secreted alkaline phosphatase, or horse-radish peroxidase.

In the methods of the present invention, a measurable signal can be detected, which results from gene expression of the reporter mRNA. For example, the signal can be selected from, but is not limited to: enzymatic activity, fluorescence, bioluminescence and combinations thereof. In one embodiment of the method to screen for RNA regulatory elements, enzymatic activity resulting from gene expression of the reporter mRNA in the presence of an RNA test sequence is measured and compared to the enzymatic activity which results from gene expression of the reporter mRNA in the absence of the RNA test sequence. In one embodiment of the methods to screen for test compounds that influence/modulate the regulatory activity of an RNA regulatory sequence, enzymatic activity resulting from gene expression of the reporter mRNA in the presence of the test compound(s) is measured and compared to the enzymatic activity in

the absence of the test compound(s). Translation of reporter mRNA and subsequent incubation with the enzyme substrate allows for the quantitative analysis of gene expression through the detection of reaction products. In preferred embodiments, the protein coding sequence of firefly luciferase is used in the reporter construct, an example  
5 of a substrate for which is luciferin, a pigment which turns over to make visible light known as bioluminescence.

### **Description of the at Least Substantially Cell-Free Translation Extract**

Another preferred embodiment of the invention features an at least substantially cell-free translation extract. Cell-free translation extracts can easily be derived by a  
10 skilled person from any source, including human, yeast, bacterial, plant and animal cells, by methods well known in the art. In preferred embodiments of the invention, the translation extract derives from human cells (for example, HeLa), yeast cells, mouse or rat cells, Chinese hamster ovary cells, *Xenopus* oocytes, reticulocytes, wheat germ, rye embryo, or bacterial cells. In a further preferred embodiment, the translation extract is a  
15 rabbit reticulocyte lysate. Rabbit reticulocyte lysates (RRL) are well known in the art and are commercially available. RRLs are highly efficient *in vitro* eukaryotic protein synthesis systems used for the translation of exogenous RNAs (either natural or generated *in vitro*). Because reticulocytes are highly specialized cells that manufacture large amounts of hemoglobin, a reticulocyte-based translation system is highly enriched for  
20 specific components of the general translation machinery. Due to the abundance of translation factors present in the RRL, the reporter mRNA used to determine reverse inhibition of translation does not require a 5' cap or a polyA tail to be translated. While RRLs are the most widely used RNA-dependent cell-free systems because of their low

background and efficient utilization of exogenous RNAs at low concentrations, the present invention is not limited to their use. However, the present invention also contemplates the use of other suitable translation systems, such as HeLa extracts, bacterial extracts, and others well known to those of skill in the art. In addition, hybrid systems  
5 such as bacterial extracts coupled with eukaryotic translation extracts may be used. Cell-free translation extracts useful for the purposes of the present invention are also commercially available, such as the rabbit reticulocyte lysate from Green Hectares Corp. (Oregon, WI), or the Hela cell extract from Promega Corp. (Madison, WI) as examples.

### **Description of the Experimental Setup**

10 The experimental setup first entails the selection of an RNA test sequence. Such a test sequence may be a synthetic sequence or a naturally occurring sequence present in a therapeutic gene of interest (or genome in the case of some pathogenic organisms), the expression of which would advantageously be modulated by small-molecule intervention at the post-transcriptional level. In the case of a sequence of a gene of interest, the RNA  
15 regulatory elements are identified and the target RNA sequence is synthesized in sufficient quantities for use in the systems of the present invention by run-off transcription (Megascript™, Ambion Inc., Austin, TX). Full-length RNA sequences of interest, as well as fragments thereof, are prepared to identify the minimal construct required for translation regulation. Subsequently, the target RNA is heat denatured and cooled to form  
20 stable secondary structure. Similarly, the synthesis of the reporter mRNA is performed by run-off transcription. The reporter construct can then be expressed in a cell-free translation extract.

In a preferred embodiment of the present invention, the inhibitory activity of the RNA test sequence or fragments thereof is first evaluated. This entails contacting the extract with each of the RNA test sequences and/or fragments to be tested, contacting the reporter with the extract (pre-treated with each of the RNA test sequences and/or fragments thereof) and selecting the minimal fragment required for efficient inhibition.

Next, the translation extract, minimal inhibitory RNA fragment, and reporter are contacted with a library of test compounds: specifically, the test compound is added to the translation extract and RNA fragment, the mixture is incubated for a set period of time, the reporter mRNA is added to the system, and the reverse inhibitors are identified by the presence of the signal produced by the translation of the reporter RNA that was previously suppressed, in the absence of the test compound, by the presence of the minimal inhibitory RNA fragments. Finally, the structure of the test compound that resulted in altered expression can be determined, leading to secondary screens with mRNA constructs that harbor the regulatory RNA sequence and, ultimately, to the development of important new compounds for molecular medicine and biotechnology.

### **Description of Compound Libraries**

Libraries screened using the methods of the present invention can comprise a variety of types of test compounds. In some embodiments, the test compounds are nucleic acid or peptide molecules. In a non-limiting example, peptide molecules can exist in a phage display library. In other embodiments, types of test compounds include, but are not limited to, peptide analogs including peptides comprising non-naturally occurring amino acids, *e.g.*, D-amino acids, phosphorous analogs of amino acids, such as  $\alpha$ -amino phosphoric acids and  $\alpha$ -amino phosphoric acids, or amino acids having non-peptide

linkages, nucleic acid analogs such as phosphorothioates and PNAs, hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose. Libraries of polypeptides or proteins can also be  
5 used.

In a preferred embodiment, the combinatorial libraries are small organic molecule libraries, such as, but not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, and diazepindiones. In another embodiment, the combinatorial libraries comprise peptoids; random bio-oligomers;  
10 diversomers such as hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; or carbohydrate libraries. Combinatorial libraries are themselves commercially available (see, *e.g.*, Advanced ChemTech Europe Ltd., Cambridgeshire, UK; ASINEX, Moscow, Russia; BioFocus plc, Sittingbourne, UK;  
15 Bionet Research (A division of Key Organics Limited), Camelford, UK; ChemBridge Corporation, San Diego, California; ChemDiv Inc, San Diego, California; ChemRx Advanced Technologies, South San Francisco, California; ComGenex Inc., Budapest, Hungary; Evotec OAI Ltd, Abingdon, UK; IF LAB Ltd., Kiev, Ukraine; Maybridge plc, Cornwall, UK; PharmaCore, Inc., North Carolina; SIDDCO Inc, Tucson, Arizona;  
20 TimTec Inc, Newark, Delaware; Tripos Receptor Research Ltd, Bude, UK; Toslab, Ekaterinburg, Russia).

In one embodiment, the combinatorial compound library for the methods of the present invention may be synthesized. There is a great interest in synthetic methods



directed toward the creation of large collections of small organic compounds, or libraries, which could be screened for pharmacological, biological or other activity (Dolle J., Comb. Chem. 3:477-517, 2001; Hall *et al.*, J. Comb. Chem. 3:125-150, 2001; Dolle J., Comb. Chem. 2:383-433, 2000; Dolle J., Comb. Chem. 1:235-282, 1999). The synthetic

5 methods applied to create vast combinatorial libraries are performed in solution or in the solid phase, *i.e.*, on a solid support. Solid-phase synthesis makes it easier to conduct multi-step reactions and to drive reactions to completion with high yields because excess reagents can be easily added and washed away after each reaction step. Solid-phase combinatorial synthesis also tends to improve isolation, purification and screening.

10 However, the more traditional solution phase chemistry supports a wider variety of organic reactions than solid-phase chemistry. Methods and strategies for the synthesis of combinatorial libraries can be found in *A Practical Guide to Combinatorial Chemistry*, A.W. Czarnik and S.H. Dewitt, eds., American Chemical Society, 1997; *The Combinatorial Index*, B.A. Bunin, Academic Press, 1998; *Organic Synthesis on Solid*

15 *Phase*, F.Z. Dörwald, Wiley-VCH, 2000; and *Solid-Phase Organic Syntheses, Vol. 1*, A.W. Czarnik, ed., Wiley Interscience, 2001.

Combinatorial compound libraries of the present invention may be synthesized using apparatuses described in US Patent No. 6,358,479 to Frisina *et al.*, U.S. Patent No. 6,190,619 to Kilcoin *et al.*, US Patent No. 6,132,686 to Gallup *et al.*, US Patent No.

20 6,126,904 to Zuellig *et al.*, US Patent No. 6,074,613 to Harness *et al.*, US Patent No. 6,054,100 to Stanchfield *et al.*, and US Patent No. 5,746,982 to Saneii *et al.* which are hereby incorporated by reference in their entirety. These patents describe synthesis apparatuses capable of holding a plurality of reaction vessels for parallel synthesis of

multiple discrete compounds or for combinatorial libraries of compounds. In one embodiment, the combinatorial compound library can be synthesized in solution. The method disclosed in U.S. Patent No. 6,194,612 to Boger *et al.*, which is hereby incorporated by reference in its entirety, features compounds useful as templates for solution phase synthesis of combinatorial libraries. The template is designed to permit reaction products to be easily purified from unreacted reactants using liquid/liquid or solid/liquid extractions. The compounds produced by combinatorial synthesis using the template will preferably be small organic molecules. Some compounds in the library may mimic the effects of non-peptides or peptides. In contrast to solid-phase synthesis of combinatorial compound libraries, liquid phase synthesis does not require the use of specialized protocols for monitoring the individual steps of a multistep solid-phase synthesis (Egner *et al.*, J.Org. Chem. 60:2652, 1995; Anderson *et al.*, J. Org. Chem. 60:2650, 1995; Fitch *et al.*, J. Org. Chem. 59:7955, 1994; Look *et al.*, J. Org. Chem. 49:7588, 1994; Metzger *et al.*, Angew. Chem., Int. Ed. Engl. 32:894, 1993; Youngquist *et al.*, Rapid Commun. Mass Spect. 8:77-81, 1994; Chu *et al.*, J. Am. Chem. Soc. 117:5419, 1995; Brummel *et al.*, Science 264:399-402, 1994; Stevanovic *et al.*, Bioorg. Med. Chem. Lett. 3:431, 1993).

Combinatorial compound libraries useful for the methods of the present invention can be synthesized on solid supports. In one embodiment, a split synthesis method, a protocol of separating and mixing solid supports during the synthesis, is used to synthesize a library of compounds on solid supports (*see* Lam *et al.*, Chem. Rev. 97:411-448, 1997; Ohlmeyer *et al.*, Proc. Natl. Acad. Sci. USA 90:10922-10926, 1993 and references cited therein). Each solid support in the final library has substantially one type

of test compound attached to its surface. Other methods for synthesizing combinatorial libraries on solid supports, wherein one product is attached to each support, will be known to those of skill in the art (see, *e.g.*, Nefzi *et al.*, Chem. Rev. 97:449-472, 1997 and US Patent No. 6,087,186 to Cargill *et al.* which are hereby incorporated by reference in their entirety). As used herein, the term "solid support" is not limited to a specific type of solid support. Rather a large number of supports are available and are known to one skilled in the art. Solid supports include silica gels, resins, derivatized plastic films, glass beads, cotton, plastic beads, polystyrene beads, alumina gels, and polysaccharides. A suitable solid support may be selected on the basis of desired end use and suitability for various synthetic protocols. For example, for peptide synthesis, a solid support can be a resin such as p methylbenzhydrylamine (pMBHA) resin (Peptides International, Louisville, KY), polystyrenes (*e.g.*, PAM-resin obtained from Bachem Inc., Torrance, CA, Peninsula Laboratories, San Carlos, CA, etc.), including chloromethylpolystyrene, hydroxymethylpolystyrene and aminomethylpolystyrene, poly (dimethylacrylamide) grafted styrene co-divinyl-benzene (*e.g.*, POLYHIPE resin, obtained from Aminotech, Ontario, Canada) polyamide resin (obtained from Peninsula Laboratories, San Carlos, CA), polystyrene resin grafted with polyethylene glycol (*e.g.*, TENTAGEL or ARGOGEL, Bayer, Tübingen, Germany) polydimethylacrylamide resin (obtained from Milligen/Bioscience, California), or Sepharose (Pharmacia, Sweden).

In one embodiment, the solid phase support is suitable for *in vivo* use, *i.e.* it can serve as a carrier or support for administration of the test compound to a patient (*e.g.*, TENTAGEL, Bayer, Tübingen, Germany). In a particular embodiment, the solid support is palatable and/or orally ingestible. In some embodiments of the present invention,

compounds can be attached to solid supports *via* linkers. Linkers can be integral and part of the solid support, or they may be nonintegral that are either synthesized on the solid support or attached thereto after synthesis. Linkers are useful not only for providing points of test compound attachment to the solid support, but also for allowing different  
5 groups of molecules to be cleaved from the solid support under different conditions, depending on the nature of the linker. For example, linkers can be, *inter alia*, electrophilically cleaved, nucleophilically cleaved, photocleavable, enzymatically cleaved, cleaved by metals, cleaved under reductive conditions or cleaved under oxidative conditions.

10 In another embodiment, the combinatorial compound libraries can be assembled *in situ* using dynamic combinatorial chemistry as described in European Patent Application 1,118,359 A1 to Lehn; Huc & Nguyen Comb. Chem. High Throughput. Screen. 4:53-74, 2001; Lehn and Eliseev, Science 291:2331-2332, 2001; Cousins *et al.*, Curr. Opin. Chem. Biol. 4: 270-279, 2000; and Karan & Miller, Drug. Disc. Today 5:67-75, 2000 which are  
15 incorporated by reference in their entirety. Dynamic combinatorial chemistry uses non-covalent interaction with a target biomolecule, including but not limited to a protein, RNA, or DNA, to favor assembly of the most tightly binding molecule that is a combination of constituent subunits present as a mixture in the presence of the biomolecule. According to the laws of thermodynamics, when a collection of molecules  
20 is able to combine and recombine at equilibrium through reversible chemical reactions in solution, molecules, preferably one molecule, that bind most tightly to a templating biomolecule will be present in greater amount than all other possible combinations. The reversible chemical reactions include, but are not limited to, imine, acyl-hydrazone,

amide, acetal, or ester formation between carbonyl-containing compounds and amines, hydrazines, or alcohols; thiol exchange between disulfides; alcohol exchange in borate esters; Diels-Alder reactions; thermal- or photoinduced sigmatropic or electrocyclic rearrangements; or Michael reactions.

5           In the preferred embodiment of this technique, the constituent components of the dynamic combinatorial compound library are allowed to combine and reach equilibrium in the absence of the target RNA and then incubated in the presence of the target RNA, preferably at physiological conditions, until a second equilibrium is reached. The second, perturbed, equilibrium (the so-called "templated mixture") can, but need not necessarily,  
10 be fixed by a further chemical transformation, including but not limited to reduction, oxidation, hydrolysis, acidification, or basification, to prevent restoration of the original equilibrium when the dynamical combinatorial compound library is separated from the target RNA. In the preferred embodiment of this technique, the predominant product or products of the templated dynamic combinatorial library can separated from the minor  
15 products and directly identified. In another embodiment, the identity of the predominant product or products can be identified by a deconvolution strategy involving preparation of derivative dynamic combinatorial libraries, as described in European Patent Application 1,118,359 A1, which is incorporated by reference in its entirety, whereby each component of the mixture is, preferably one-by-one but possibly group-wise, left out of the mixture  
20 and the ability of the derivative library mixture at chemical equilibrium to bind the target RNA is measured. The components whose removal most greatly reduces the ability of the derivative dynamic combinatorial library to bind the target RNA are likely the

components of the predominant product or products in the original dynamic combinatorial library.

### Description of Methods for Determining the Structure of the Test Compound

If the library comprises arrays or microarrays of compounds, wherein each  
5 compound has an address or identifier, the compound can be deconvoluted, *e.g.*, by cross-referencing the positive sample to original compound list that was applied to the individual test assays. If the library is a peptide or nucleic acid library, the sequence of the compound can be determined by direct sequencing of the peptide or nucleic acid. Such methods are well known to one of skill in the art. A number of physico-chemical  
10 techniques can be used for the *de novo* characterization of compounds bound to the target RNA. Examples of such techniques include, but are not limited to, mass spectrometry, NMR spectroscopy, X-ray crystallography and vibrational spectroscopy. The characterization of compounds bound to the target RNA allows for the identification of active molecules from mixtures obtained from combinatorial chemistry libraries.

#### 15 Mass Spectrometry

Mass spectrometry (*e.g.*, electrospray ionization (“ESI”), matrix-assisted laser desorption-ionization (“MALDI”), and Fourier-transform ion cyclotron resonance (“FT - ICR”) can be used for elucidating the structure of a compound. MALDI uses a pulsed  
20 laser for desorption of the ions and a time-of-flight analyzer, and has been used for the detection of noncovalent tRNA:amino-acyl-tRNA synthetase complexes (Gruic-Sovulj *et al.*, J. Biol. Chem. 272:32084-32091, 1997). However, covalent cross linking between the target nucleic acid and the compound is required for detection, since a non-covalently

bound complex may dissociate during the MALDI process. ESI mass spectrometry ("ESI-MS") has been of greater utility for studying non-covalent molecular interactions because, unlike the MALDI process, ESI-MS generates molecular ions with little to no fragmentation (Xavier *et al.*, Trends Biotechnol. 18(8):349-356, 2000). ESI MS has been used to study the complexes formed by HIV Tat peptide and protein with the TAR RNA (Sannes-Lowery *et al.*, Anal. Chem. 69:5130-5135, 1997). Fourier-transform ion cyclotron resonance ("FT-ICR") mass spectrometry provides high-resolution spectra, isotope-resolved precursor ion selection, and accurate mass assignments (Xavier *et al.*, Trends Biotechnol. 18(8):349-356, 2000). FT-ICR has been used to study the interaction of aminoglycoside antibiotics with cognate and non-cognate RNAs (Hofstadler *et al.*, Anal. Chem. 71:3436-3440, 1999; and Griffey *et al.*, Proc. Natl. Acad. Sci. USA 96:10129-10133, 1999). As true for all of the mass spectrometry methods discussed herein, FT-ICR does not require labeling of the target RNA or a compound. An advantage of mass spectroscopy is not only the elucidation of the structure of the compound, but also the determination of the structure of the compound bound to a target RNA. Such information can enable the discovery of a consensus structure of a compound that specifically binds to a target RNA.

### NMR Spectroscopy

NMR spectroscopy is a valuable technique for identifying complexed target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects, and NMR-based approaches have been used in the identification of small molecule binders of protein drug targets (Xavier *et al.*, Trends Biotechnol. 18(8):349-356, 2000). The determination of structure-activity

relationships (“SAR”) by NMR is the first method for NMR described in which small molecules that bind adjacent subsites are identified by two-dimensional <sup>1</sup>H-<sup>15</sup>N spectra of the target protein (Shuker *et al.*, Science 274:1531-1534, 1996). The signal from the bound molecule is monitored by employing line broadening, transferred NOEs and pulsed  
5 field gradient diffusion measurements (Moore, Curr. Opin. Biotechnol. 10:54-58, 1999).

A strategy for lead generation by NMR using a library of small molecules has been recently described (Fejzo *et al.*, Chem. Biol. 6:755-769, 1999). In one embodiment of the present invention, the target nucleic acid complexed to a compound can be determined by SAR by NMR. Furthermore, SAR by NMR can also be used to elucidate the structure of a  
10 compound. As described above, NMR spectroscopy is a technique for identifying binding sites in target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects.

Examples of NMR that can be used for the invention include, but are not limited to, one-dimensional NMR, two-dimensional NMR, correlation spectroscopy (“COSY”),  
15 and nuclear Overhauser effect (“NOE”) spectroscopy. Such methods of structure determination of compounds are well-known to one of skill in the art. Similar to mass spectroscopy, an advantage of NMR is the not only the elucidation of the structure of the compound, but also the determination of the structure of the compound bound to the target RNA. Such information can enable the discovery of a consensus structure of a  
20 compound that specifically binds to a target RNA.



## X-ray Crystallography

X-ray crystallography can be used to elucidate the structure of a compound. For a review of x-ray crystallography see, *e.g.*, Blundell *et al.*, Nat Rev Drug Discov 1(1):45-54, 2002. The first step in x-ray crystallography is the formation of crystals. The formation of  
5 crystals begins with the preparation of highly purified and soluble samples. The conditions for crystallization are then determined by optimizing several solution variables known to induce nucleation, such as pH, ionic strength, temperature, and specific concentrations of organic additives, salts and detergent. Techniques for automating the crystallization process have been developed for the production of high-quality protein  
10 crystals. Once crystals have been formed, the crystals are harvested and prepared for data collection. The crystals are then analyzed by diffraction (such as multi-circle diffractometers, high-speed CCD detectors, and detector off-set). Generally, multiple crystals must be screened for structure determinations.

## Vibrational Spectroscopy

15 Vibrational spectroscopy (*e.g.* infrared (IR) spectroscopy or Raman spectroscopy) can be used for elucidating the structure of a compound. Infrared spectroscopy measures the frequencies of infrared light (wavelengths from 100 to 10,000 nm) absorbed by the compound as a result of excitation of vibrational modes according to quantum mechanical selection rules which require that absorption of light cause a change in the electric dipole  
20 moment of the molecule. The infrared spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound. Infrared spectra can be measured in a scanning

mode by measuring the absorption of individual frequencies of light, produced by a grating which separates frequencies from a mixed frequency infrared light source, by the compound relative to a standard intensity (double-beam instrument) or pre-measured ('blank') intensity (single-beam instrument).

5           In a preferred embodiment, infrared spectra are measured in a pulsed mode ("FT-IR") where a mixed beam, produced by an interferometer, of all infrared light frequencies is passed through or reflected off the compound. The resulting interferogram, which may or may not be added with the resulting interferograms from subsequent pulses to increase the signal strength while averaging random noise in the electronic signal, is  
10   mathematically transformed into a spectrum using Fourier Transform or Fast Fourier Transform algorithms.

          Raman spectroscopy measures the difference in frequency due to absorption of infrared frequencies of scattered visible or ultraviolet light relative to the incident beam. The incident monochromatic light beam, usually a single laser frequency, is not truly  
15   absorbed by the compound but interacts with the electric field transiently. Most of the light scattered off the sample will be unchanged (Rayleigh scattering) but a portion of the scatter light will have frequencies that are the sum or difference of the incident and molecular vibrational frequencies. The selection rules for Raman (inelastic) scattering require a change in polarizability of the molecule.

20           While some vibrational transitions are observable in both infrared and Raman spectrometry, most are observable only with one or the other technique. The Raman spectrum of any molecule is a unique pattern of absorption wavelengths of varying

intensity that can be considered as a molecular fingerprint to identify any compound.

Raman spectra are measured by submitting monochromatic light to the sample, either passed through or preferably reflected off, filtering the Rayleigh scattered light, and detecting the frequency of the Raman scattered light. An improved Raman spectrometer is

5 described in US Patent No. 5,786,893 to Fink *et al.*, which is hereby incorporated by reference. Vibrational microscopy can be measured in a spatially resolved fashion to address single beads by integration of a visible microscope and spectrometer. A microscopic infrared spectrometer is described in U.S. Patent No. 5,581,085 to Reffner *et al.*, which is hereby incorporated by reference in its entirety. An instrument that

10 simultaneously performs a microscopic infrared and microscopic Raman analysis on a sample is described in U.S. Patent No. 5,841,139 to Sostek *et al.*, which is hereby incorporated by reference in its entirety.

In one embodiment of the method, compounds are synthesized on polystyrene beads doped with chemically modified styrene monomers such that each resulting bead

15 has a characteristic pattern of absorption lines in the vibrational (IR or Raman) spectrum, by methods including but not limited to those described by Fenniri *et al.*, J. Am. Chem. Soc. 123:8151-8152, 2000. Using methods of split-pool synthesis familiar to one of skill in the art, the library of compounds is prepared so that the spectroscopic pattern of the bead identifies one of the components of the compound on the bead. Beads that have been

20 separated according to their ability to bind target RNA can be identified by their vibrational spectrum. In one embodiment of the method, appropriate sorting and binning of the beads during synthesis then allows identification of one or more further components of the compound on any one bead. In another embodiment of the method,

partial identification of the compound on a bead is possible through use of the spectroscopic pattern of the bead with or without the aid of further sorting during synthesis, followed by partial resynthesis of the possible compounds aided by doped beads and appropriate sorting during synthesis. In another embodiment, the IR or Raman spectra of compounds are examined while the compound is still on a bead, preferably, or after cleavage from bead, using methods including but not limited to photochemical, acid, or heat treatment. The compound can be identified by comparison of the IR or Raman spectral pattern to spectra previously acquired for each compound in the combinatorial library.

## EXAMPLES

The following examples are provided to better illustrate the claimed invention and are not to be interpreted as limiting the scope of the invention. To the extent that specific materials are mentioned, it is merely for purposes of illustration and is not intended to limit the invention. Unless otherwise specified, general cloning procedures are used, such as those set forth in the following: Sambrook *et al.*, Molecular Cloning, Cold Spring Harbor Laboratory, 2001, Ausubel *et al.* (eds.); and Current Protocols in Molecular Biology, John Wiley & Sons, 2000. One skilled in the art may develop equivalent means or reactants without departing from the scope of the invention.

### EXAMPLE 1

#### Construction of the Reporter mRNA

This example illustrates the construction of the reporter mRNA to determine the level of reverse inhibition of translation. The reporter plasmid pT7 is constructed to contain the following elements: T7 promoter sequence and multiple cloning site (MCS).

The non-natural MCS sequences used for the initial insertion of the Luciferase open reading frame (ORF) can also be used for generating fusion proteins. The plasmid pT7-luc contains the protein coding sequence of firefly luciferase cloned downstream of the T7 promoter sequence at the NcoI and BglII restriction endonuclease sites. Figure 4 shows the components of an exemplary reporter DNA construct used to produce reporter mRNA for use in the methods of the present invention, a description for which is in the Brief Description of the Drawings. pT7-luc was digested with BamHI, which cuts downstream of the firefly luciferase coding sequence, to generate the linear DNA template used for run-off transcription. Run-off transcription was performed using the Megascript kit (Ambion Inc., Austin, TX; performed according to the manufacturers instructions) to prepare the reporter mRNA. Accordingly, there is no UTR regulatory element present.

Equivalent reporter mRNAs (besides luciferase) include, but are not limited to, the ORFs of renilla luciferase, click beetle luciferase, green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent protein (RFP), cyan fluorescent protein (CFP), blue fluorescent protein (BFP), *beta*-galactosidase, *beta*-glucuronidase, *beta*-lactamase, chloramphenicol acetyltransferase (CAT), secreted alkaline phosphatase (SEAP), or horse-radish peroxidase (HRP).

## EXAMPLE 2

### Determining the Inhibitory Activity of an RNA Test Sequence

This example provides a detailed protocol for determining the inhibitory activity of an RNA test sequence derived from the UTR of HCV, which contains a well-characterized IRES RNA regulatory element.

**Materials**

Folding buffer B [50 mM TRIS (pH-7.5), 2.5 mM MgCl<sub>2</sub>, 10 mM KCl, 250 mM NaCl].

mRNA (Firefly Luciferase; non-capped; transcribed with T7 polymerase).

- 5 RNA inhibitor (HCV IRESIII RNA fragment; transcribed with T7 polymerase; 10 µM, folded in buffer B at 95°C for 5 min and cooled in ice for 10 min).

Rabbit Reticulocyte Lysate [RRL; Green Hectares Inc., Oregon, WI; prepared by suspending settled Red Blood Cells in equal volume of water and supplemented with Hemin (0.013 mg/ml), creatine kinase (0.05 mg/ml), and tRNA (0.125 mg/ml)].

- 10 5X Reaction buffer A [0.15 mM of each amino acid mix, 500 mM KOAc, 2.5 mM Mg(OAc)<sub>2</sub> and 50 mM creatine phosphate].

DMSO (dimethylsulfoxide).

Nuclease free water.

15 **Methods**

1. **Construction of Inhibitory RNA fragments:**

The plasmid pGEM3-HCV2b-IRES contains sequence coding for nucleotides 18 to 356 (SEQ ID NO: 1) of HCV-IRES RNA (genotype 2b) inserted at the BamHI restriction endonuclease site of the cloning vector pGEM-3 (Promega Corp., Madison, WI; Figure 2A). RNA secondary structure prediction algorithms, as well as biochemical

20

and genetic data, support a representation of this sequence as shown in Figure 2B.

Domains II-IV of the HCV-IRES have been reported to be necessary and sufficient to facilitate internal ribosome entry through recruitment of translational initiation factors.

Thus, this RNA sequence, as well as fragments thereof, was used to inhibit translation of

5 luciferase enzyme from reporter mRNA. To generate RNA fragments, primers complementary to sequences between structured regions were prepared, with 5' forward primers designed to place a T7 promoter sequence (taatacgactcactatagg) upstream of the sequence of interest for use in run-off transcription reactions. For fragments less than 50

nucleotides, run-off transcription was performed directly from annealed primers. For

10 sequences greater than 50 nucleotides, the primers were used to PCR amplify the region of interest, the PCR products were TA cloned into the plasmid pCR4-TOPO (Invitrogen Corp., Carlsbad, CA; performed according to the manufacturers instructions), the cloned fragment isolated through restriction enzyme digestion using EcoRI, and run-off

transcription was performed directly from the purified fragments. The RNA fragments, as

15 well as the method in which they were synthesized and the sequences of the primers used for their synthesis, are listed below:

**Direct annealing:**

HCV-IIb (SEQ ID NO: 2)

20

HCV-IIb:

taatacgactcactataggcagaaagcgtctagccatggcgtagtatga (SEQ ID NO: 3)

HCV-IIb-RC: tcatactaacgccatggctagacgctttctg cctatagtgagtcgtatta  
(SEQ ID NO: 4)

25

HCV-IV (SEQ ID NO: 5)

HCV-IV: taatacgactcactata ggaccgtgcatcatgagcacgaatcc (SEQ ID NO: 6)

HCV-IV-RC: ggattcgtgctcatgatgcacggctcctatagtgagtcgtatta (SEQ ID NO: 7)

3. HCV-IIIb (SEQ ID NO: 8)

5 HCV-IIIb: taatacgactcactataggccggaagactgggtcctttcttgataaaccactctatgtccgg  
(SEQ ID NO: 9)

HCV-IIIb-RC: ccggacatagagtgggtttatccaagaaaggacccagtctttccggcctatagtgagtcgtatta  
(SEQ ID NO: 10)

4. HCV-IIa (SEQ ID NO: 11)

10 HCV-IIa: taatacgactcactataggcctgtgaggaaactactgtcttcacttcggtgtcgtacagcctccagg  
(SEQ ID NO: 12)

15 HCV-IIa-RC:  
cctggaggctgtacgacaccgaagtgaagacagtagttcctcacaggcctatagtgagtcgtatta  
(SEQ ID NO: 13)

#### PCR Primers:

1. HCV-IIab (SEQ ID NO: 14)

HCV-IIab-For: taatacgactcactataggcgacactccgccatgagtcac (SEQ ID NO: 15)

20 HCV-IIab-Back: ggcttgaggctgtacgacactcatac (SEQ ID NO: 16)

2. HCV-IIIabc (SEQ ID NO: 17)

HCV-IIIabc-For: taatacgactcactataggtagtgtctgcggaaccgg (SEQ ID NO: 18)

HCV-IIIabc-Back: tagcagtcttgcgggggcacg (SEQ ID NO: 19)

3. HCV-IIIabcd (SEQ ID NO: 20)

25 HCV-IIIabcd-For: taatacgactcactataggagagccatagtggctctgcgg (SEQ ID NO: 21)

HCV-IIIabcd-Back: agtaccacaaggcctttcgc (SEQ ID NO: 22)

4. HCV-IIIdef (SEQ ID NO: 23)

HCV-IIIdef-For: taatacgactcactataggctgtagccgagtagcg (SEQ ID NO: 24)



- HCV-IIIdef-Back: tacgagacctcccggggcac (SEQ ID NO: 25)
5. HCV-III (SEQ ID NO: 26)
- HCV-III-For: Taatacgactcactataggccccccctcccgaggagacc (SEQ ID NO: 27)
- HCV-IIIdef-Back: tacgagacctcccggggcac (SEQ ID NO: 28)
- 5 6. HCV-II-IIIb (SEQ ID NO: 29)
- HCV-IIab-For: taatacgactcactataggcgacactccgccatgagtcac (SEQ ID NO: 30)
- HCV-IIIb-Back: ggaatgaccggacatagagtgggtttatc (SEQ ID NO: 31)
7. HCV-II-IIIabc (SEQ ID NO: 32)
- HCV-IIab-For: taatacgactcactataggcgacactccgccatgagtcac (SEQ ID NO: 33)
- 10 HCV-IIIabc-Back: tagcagtcttgcgggggcacg (SEQ ID NO: 34)
8. HCV-II-IIIabcd (SEQ ID NO: 35)
- HCV-IIab-For: taatacgactcactataggcgacactccgccatgagtcac (SEQ ID NO: 36)
- HCV-IIIabcd-Back: agtaccacaaggcctttcgc (SEQ ID NO: 37)
9. HCV-II-III (SEQ ID NO: 38)
- 15 HCV-IIab-For: taatacgactcactataggcgacactccgccatgagtcac (SEQ ID NO: 39)
- HCV-IIIdef-Back: tacgagacctcccggggcac (SEQ ID NO: 40)
10. HCV-III/IV (SEQ ID NO: 41)
- HCV-III-For:
- 20 taatacgactcactataggccccccctcccgaggagacc (SEQ ID NO: 42)
- HCV-IV-RC:
- ggattcgtgctcatgatgcacggtcctatagtgcgtgattc (SEQ ID NO: 43)
11. HCV-II/III/IV (SEQ ID NO: 44)
- 25 HCV-IIab-For: taatacgactcactataggcgacactccgccatgagtcac (SEQ ID NO: 45)
- HCV-IV-RC:
- ggattcgtgctcatgatgcacggtcctatagtgcgtgattc (SEQ ID NO: 46)
- 30

## 2. Translation reaction with inhibitor RNA:

Translation reactions were performed *in vitro* to determine the inhibitory effect of HCV IRESIII RNA and fragments thereof and contained (20  $\mu$ l): 4  $\mu$ l pre-treated RRL, 4  $\mu$ l 5X Reaction Buffer A, 5  $\mu$ l of HCV IRES RNA (1  $\mu$ M in 20% B), 5  $\mu$ l mRNA (40 ng/ $\mu$ l), and 2  $\mu$ l 10% DMSO. A control reaction was complemented with 5  $\mu$ l 20% B instead of IRESIII and was considered for 100% activity. The translation reaction mixture was pre-incubated for 30 min at room temperature in the absence of mRNA. Translation reaction was started by addition of mRNA and incubated at 30°C for 60 min. Translation reaction was stopped and Luciferase activity was measured using ViewLux™ (Perkin-Elmer Scientific Instruments, Boston, MA; detector set at 10 sec measurement and 6X binning; Figure 3) upon the addition of 50  $\mu$ l LucLite –Plus™ luciferase assay reagent (Promega Corp., Madison, WI). Based upon this analysis, RNA fragment III was chosen for further study to identify test compounds that reverse inhibition.

## EXAMPLE 4

### Determining the Reverse Inhibition Activity of Test Compound.

This example provides a detailed protocol for determining the reverse inhibition activity of a test compound (PTC-0099870) that was previously determined to interact with HCV-IRES RNA and inhibit internal ribosome entry.

#### Materials

1. Folding buffer B [50 mM TRIS (pH-7.5), 2.5 mM MgCl<sub>2</sub>, 10 mM KCl, 250 mM NaCl]

2. mRNA (Firefly Luciferase; non-capped; transcribed with T7 polymerase).
3. RNA inhibitor (HCV IRESIII RNA fragment; transcribed with T7 polymerase; 10  $\mu$ M, folded in buffer B at 95°C for 5 min and cooled in ice for 10 min).
- 5 4. Rabbit Reticulocyte Lysate [RRL; Green Hectares, Inc., Oregon, WI; prepared by suspending settled Red Blood Cells in equal volume of water and supplemented with Hemin (0.013 mg/ml), creatine kinase (0.05 mg/ml), and tRNA (0.125 mg/ml)]
5. 5X Reaction buffer A [0.15 mM of each amino acid mix, 500 mM KOAc, 2.5 mM Mg(OAc)<sub>2</sub> and 50 mM creatine phosphate]
- 10 DMSO (dimethylsulfoxide)
- Test compound (PTC-0099870; dissolved in 10% DMSO)
8. Nuclease free water

## 15 **Methods**

### 1. **Translation reaction with inhibitor RNA and test compound:**

Translation reactions were performed *in vitro* to detect the reverse inhibition by test compound and contained (20  $\mu$ l): 4  $\mu$ l pre-treated RRL, 4  $\mu$ l 5X Reaction Buffer A, 5  $\mu$ l of IRESIII (400 nM in 20% B), 5  $\mu$ l mRNA (40 ng/ $\mu$ l), and 2  $\mu$ l of compounds

(dissolved in 10% DMSO). A control reaction was complemented with 5  $\mu$ l 20% B instead of IRESIII and 2  $\mu$ l 10% DMSO instead of compound and was considered for 100% activity. The translation reaction mixture was pre-incubated for 30 min at room temperature in the absence of mRNA. Translation reaction was started by addition of mRNA and incubated at 30°C for 60 min. Translation reaction was stopped and Luciferase activity was measured using ViewLux™ (Perkin-Elmer Scientific Instruments, Boston, MA; detector set at 10 sec measurement and 6X binning; Figure 5) upon the addition of 50  $\mu$ l LucLite –Plus™ luciferase assay reagent (Promega Corp., Madison, WI).

## EXAMPLE 5

### High-Throughput Screening for Compounds that Reverse Inhibition

This example describes how the systems and methods of the present invention are scaled-up and adapted for use in a high-throughput screening assay. The assay is conducted in 384-well format. The chemicals in library are screened at 15  $\mu$ M.

#### Materials

1. Folding buffer B [50 mM TRIS (pH-7.5), 2.5 mM  $MgCl_2$ , 10 mM KCl, 250 mM NaCl]
2. mRNA (Firefly Luciferase; non-capped; transcribed with T7 polymerase).
3. RNA inhibitor (HCV IRESIII RNA fragment; transcribed with T7 polymerase; 10  $\mu$ M, folded in buffer B at 95°C for 5 min and cooled in ice for 10 min).
4. Rabbit Reticulocyte Lysate [RRL; Green Hectares Inc., Oregon, WI;

prepared by suspending settled Red Blood Cells in equal volume of water and supplemented with Hemin (0.013 mg/ml), creatine kinase (0.05 mg/ml), and tRNA (0.125 mg/ml)]

5. 5X Reaction buffer A [0.15 mM of each amino acid mix, 500 mM KOAc, 2.5 mM Mg(OAc)<sub>2</sub> and 50 mM creatine phosphate]
6. DMSO (dimethylsulfoxide)
7. Compound library (dissolved in 10% DMSO; one compound/well format; 386 well plate)
8. Nuclease treated (free) water (NT H<sub>2</sub>O)

## 10 Methods

### 1. Preparation of the mono-cistronic luciferase reporter construct (mRNA)

The reporter plasmid pT7-luc is linearized by digestion with BamHI to generate the double-stranded linear DNA template used for run-off transcription. Run-off transcription is performed using the Megascript kit (Ambion, Austin, TX) to generate the reporter mRNA for use in the high-throughput screening assay. 10X reaction buffer is prepared by combining 100  $\mu$ l 1M MgCl<sub>2</sub>, 400  $\mu$ l 1M Tris pH 7.5, 100  $\mu$ l 1M NaCl, 6.7  $\mu$ l 3M spermadine, 100  $\mu$ l (2.5 u/ $\mu$ l) pyrophosphatase, 200  $\mu$ l 1M DTT, and 93.3  $\mu$ l NT H<sub>2</sub>O, bringing the total volume to 1 ml and the final concentration of the reagents to the respective values listed in Table 1 below:

**Table 1**

Reagent	Vol (1ml)	Final Conc.
MgCL <sub>2</sub> (1M)	100 $\mu$ l	100 mM
Tris Ph 7.5 (1M)	400 $\mu$ l	400 mM
NaCl (1M)	100 $\mu$ l	100 mM
Spermadine (3M)	6.7 $\mu$ l	20 mM
Pyrophosphatase (2.5 $\mu$ /ml)	100 $\mu$ l	25 units/ml
DTT (1M)	200 $\mu$ l	200 mM
NT H <sub>2</sub> O	93.3 $\mu$ l	QS 1ml

- The 10 x 1 ml transcription reaction is set up by combining the double-stranded DNA template of the reporter with a mixture of nucleotides and T7 polymerase in
- 5 Reaction Buffer in a final volume of 1 ml in NT H<sub>2</sub>O. The reaction was incubated at 37°C for 4 hours. The final concentrations of the reagents are listed in Table 2 below:

**Table 2**

Reagent	Vol	Final Conc.
dsDNA (1mg/ml)	100 $\mu$ l	0.1 mg/ml
NTPs (75mM)	268 $\mu$ l	5 mM
10X Reaction Buffer	100 $\mu$ l	1X
T7 Polymerase Plus		2,000 $\mu$ /ml
NT H <sub>2</sub> O		QS 1 ml

- 10 Next, the sample is treated with DNaseI at a final concentration of 2  $\mu$ /ml to eliminate any contaminating residual DNA, and the transcription product (RNA) is precipitated with 700  $\mu$ L 7.5 M LiCl/50 mM EDTA in a final volume of 1 ml, incubated

for 2 hours at  $-20^{\circ}\text{C}$ , spun for 30 min at  $4^{\circ}\text{C}$  at a speed for 14,000 rpm. The pellet is washed twice with 70% ethanol and dried by inversion. After drying, the pellet is resuspended by adding 1 ml water and vortexing. The yield of a typical 10 ml transcription reaction is approximately 50 mg RNA. The concentration of the RNA is measured by taking the OD at 260 nm on a CARY-100BIO spectrophotometer (Varion Corp., Palo Alto, CA). Concentration is calculated by the following formula: RNA ( $\mu\text{g/ml}$ ) =  $\text{OD}_{260} \times 40 \times 20$  (dilution factor). The RNA is diluted to a concentration of 40 ng/ $\mu\text{l}$  in NT H<sub>2</sub>O.

## 10    2.    **Preparation of the Inhibitory RNA construct**

The linear template dsDNA of the RNA test fragment, here the IRESIII domain, is constructed as described for the reporter construct above, with the plasmid containing the T7 promoter sequence and multiple cloning site (MCS) for the insertion of the coding sequence of the RNA element. After insertion of the coding sequence of the RNA element, the fragment containing the IRESIII domain is purified by digestion with the EcoRI restriction enzyme to generate the double-stranded linear DNA template used for run-off transcription. Run-off transcription is performed using the Megascript kit (Ambion, Austin, TX); a 20 x 1ml transcription reaction is set up using the Megascript T7 kit with the following reagents, as shown in Table 3 below:

20

**Table 3**

Reagent	Vol	Final Conc.
dsDNA (0.1mg/ml)	75 $\mu$ l	100 nM (100 pmoles)
NTPs (75mM)	268 $\mu$ l	5 mM
10X Reaction Buffer	100 $\mu$ l	1X
T7 Polymerase Plus	100 $\mu$ l	2,000 $\mu$ /ml
NT H <sub>2</sub> O		QS 1 ml

The reaction is incubated at 37°C for 4 hours and subsequently treated with Dnase I at a final concentration of 2 u/ $\mu$ l. The RNA product of the transcription reaction is

5 precipitated with LiCl as described previously, and incubated for 2 hours at – 20°C, spun for 30 min at 4°C at a speed for 14,000 rpm. The pellet is washed twice with 70% ethanol and dried by inversion. After drying, the pellet is resuspended by adding 1 ml water and vortexing. The concentration of the RNA element is measured as described above. For the RNA folding reaction, Folding Buffer B is prepared by combining the reagents listed

10 in Table 4 below:

**Table 4**

Reagent	Vol	Final Conc.
TRIS (pH 7.5) (1M)	25.00 ml	50 mM
MgCl <sub>2</sub> (1M)	1.25 ml	2.5 mM
KCl (1M)	5.00 ml	10 mM
NaCl (4M)	31.25 ml	250 mM
NT H <sub>2</sub> O	437.5 ml	
Vf =	500 ml	



The RNA element is diluted to 10  $\mu$ M in Folding Buffer B, heated to 90°C for 5 min and cooled on ice for 15 min. The folded RNA element is then diluted five-fold to 2  $\mu$ M in 100% Buffer B and re-diluted 5-fold in NT H<sub>2</sub>O to a final concentration of 400 nM. The final concentration of B Buffer is 20%.

5

### 3. Preparation of the 5X Reaction buffer

5X Reaction buffer is prepared by combining the reagents in Table 5 below at the concentrations listed:

**Table 5**

Reagent	Vol	Final Conc.
Amino Acids (2mM)	112.5 ml	0.15 mM
KOAc (5M)	750 ml	500 mM
Mg(OAc) <sub>2</sub> (2.5M)	3.75 ml	2.5 mM
Creatine Phosphate (0.2M)	375 ml	50 mM
NT H <sub>2</sub> O		QS 1.5 L

10

### 4. Preparation of 96-well Master Standard Plate

For the high-throughput screen, each of the 384-well assay plates contains standard reactions as internal controls, including (i) the experimental “low” signal, with reactions performed in the absence of positive control test compound, (ii) the experimental “high” signal, with reactions performed in the presence of a high concentration of positive control test compound (PTC-0099870), and (iii) the standard curve signal, with reactions performed in the presence of a titration of PTC-0099870. A 96-well PolyPro deep-well plate is used to generate the Master Standard Plate, with each well containing standards @ 10X concentration. As shown in Table 6 below, column #1,

15

rows a-d, contains the experimental low signal (10% DMSO); column #1, rows e-h, contains the experimental high signal (PTC-0099870 in 10% DMSO); column #2, rows a-h, contains the standard curve signal (0-28uM PTC-0099870 in 10% DMSO):

5

**Table 6**

	1	2
<b>A</b>	10% DMSO	300 $\mu$ M PTC-0099870
<b>B</b>	10% DMSO	250 $\mu$ M PTC-0099870
<b>C</b>	10% DMSO	200 $\mu$ M PTC-0099870
<b>D</b>	10% DMSO	150 $\mu$ M PTC-0099870
<b>E</b>	300 $\mu$ M PTC-0099870	100 $\mu$ M PTC-0099870
<b>F</b>	300 $\mu$ M PTC-0099870	50 $\mu$ M PTC-0099870
<b>G</b>	300 $\mu$ M PTC-0099870	25 $\mu$ M PTC-0099870
<b>H</b>	300 $\mu$ M PTC-0099870	0uM PTC-0099870

Each well of the 96-well Master Standard Plate will be used to robotically aliquot 2  $\mu$ l of standards to each of 4 wells of the 384-well assay plate; thus, each of the 96-well Master Standard Plates will be used to generate 25 384-well assay plates (240  $\mu$ L/well), and columns 1-4 of each assay plate will contain standards as internal controls. 4 Master Standard Plates are prepared each day, resulting in an experimental volume of 100 384-well assay plates/day.

#### 5. Preparation of Master Translation Reaction Mix Plate

15

The translation reaction mix contains RRL and IRESIII RNA element for pre-incubation with test compounds in the 384-well assay plate. A 384-well PolyPro deep-

well plate is used to generate the Master Translation Reaction Mix Plate, with each well containing a predetermined amount of translation reaction mix. Each well of the 384-well Master Translation Reaction Mix Plate will be used to robotically aliquot 13  $\mu$ l of standards to each well of the 384-well assay plate; thus, each of the 384-well Master Translation Reaction Mix Plates will be used to generate 10 384-well assay plates (125  $\mu$ L/well). 10 Master Translation Reaction Mix Plates are prepared each day, resulting in an experimental volume of 100 384-well assay plates/day. Preparation of the Master Translation Reaction Mix is shown in Table 7 below.

**Table 7**

Reagent (no DMSO)	Vol (1X)	Vol (42000X)	Final Conc.
RRL (100%)	3 $\mu$ l	168.0 ml	20%
Reaction Buffer (5X)	5 $\mu$ l	168.0 ml	1X
IRESIII RNA (400 nM)	5 $\mu$ l	210.0 ml	400 nM
DMSO (4%)			1%
Vf =	13 $\mu$ l	546.0 ml	

## 6. Preparation of Master Assay Plates

High-throughput screening robotics is employed at this point to prepare the Master Assay Plates. First, 2  $\mu$ L of standards in 10% DMSO (from Master Standard Plate; see above) are added to columns 1-4 of each assay plate, and 2  $\mu$ L of each test compound in 10% DMSO (from compound library arrayed in 384-well plate format at 10X concentration; columns 5-24) is added to each of the remaining wells. Next, 13  $\mu$ L of translation reaction mix (from Master Translation Reaction Mix Plates; see above) are added to each well of the Master assay Plate and incubated at room temperature for 30

minutes to allow reaction components to reach equilibrium. Next, 10  $\mu$ L of mRNA are added to each well of the Master assay Plate and incubated at 30°C for 30 minutes to allow for gene expression (translation).